
Title: Codebook - MABco Analysis (Case Study 1)

Author: Data Busters Inc. (K Henley, G Lane, J McDaniel)

Date: 10/22/2017

Output: pdf

****Project Description****

The client has provided data on key craft beer offerings in the U.S., as well as the corresponding breweries where the beers are produced. The data is being analyzed to produce findings that may help the client gain a better understanding of the kinds of craft beer being offered, as well as the locations where they are produced and made available.

****Study design and data processing****

We have two data sets ("beers" and "breweries"). We are conducting a multivariate analysis of the data provided, and in total have (10) unique variables to evaluate.

****Collection of the raw data****

The data sets were provided to DBI by the Mega Awesome Beer Co. (MABco), but the actual collection methodology and sourcing is unknown.

****Notes on the original (raw) data****

Raw data in each file contains headers/variable names. Our "beers" data set contains (7) variables that are Name, Beer_ID, ABV, IBU, Brewery_id, Style, Ounces. Our "breweries" data set contains (4) variables that are Brew_ID, Name, City, State. Of note, we have a shared or linked variable (Brewery_id & Brew_ID) that allows us to cross reference the two data sets.

****Guide to create the tidy data file****

1) Download/Import data for "beers" and "breweries"

2) Merge the two files based on the shared variable (Brewery_id & Brew_ID)

3) Assign the following variable names to the merged data set: BreweryID, Name, BeerID, ABV, IBU.

4) Convert columns incorrectly listed as ordered factors (ABV, IBU, Ounces) to numeric.

5) Sort the merged data set by state and city.

****Description of the variables in the tidy "merged" file****

#Data Frame with 2410 observations and 10 variables#

1) BreweryID - integer *This variable provides an assigned identification number for each brewery.

2) Name - Factor *This variable lists the names assigned to each unique beer offering. (2305 levels)

3) BeerID - integer *This variable provides an assigned identification number for each beer.

4) ABV - numeric *This variable provides a numeric value assigned to each beer's alcohol by volume (ABV).

5) IBU - integer *This variable provides a numeric value assigned to each beer's bitterness rating.

6) Style - Factor *This variable lists the names of the style of beer assigned to each offering. (100 levels)

7) Ounces - numeric *This variable provides a numeric value assigned to the number of ounces contained in each bottle of beer.

8) BreweryName - Factor *This variable lists the names assigned to each unique brewery. (551 levels)

9) City - Factor *This variable lists the name of each city where each brewery is located. (384 levels)

10) State - Factor *This variable lists the name of each state where each brewery is located. (51 levels)

****Sources****

Data Sets provided for analysis by Mega Awesome Beer Co. (MABco) for download via the following URL:

https://2ds.datascience.smu.edu/local/files/index.php?course_id=104&group=290&userid=1075&groupid=-1